

Variance Component Testing in Generalized Linear Mixed Models for Longitudinal/Clustered Data and Other Related Topics

Daowen Zhang^{1,*}, Xihong Lin²

¹ Department of Statistics, North Carolina State University, Raleigh, NC 27695

² Department of Biostatistics, Harvard School of Public Health, Boston, MA 02115

**email*: zhang@stat.ncsu.edu

KEY WORDS: Boundary parameters; Likelihood ratio tests; Mixtures of chi-squares;
Penalized splines, Score tests, Smoothing splines

1 Introduction

Linear mixed models (Laird and Ware, 1982) and generalized linear mixed models (GLMMs) (Breslow and Clayton, 1993) have been widely used in many research areas, especially in the area of biomedical research, to analyze longitudinal and clustered data and multiple outcome data. In a mixed effects model, subject-specific random effects are used to explicitly model between-subject variation in the data and often assumed to follow a mean zero parametric distribution, e.g., multivariate normal, that depends on some unknown variance components. A large literature was developed in the last two decades for the estimation of regression coefficients and variance components in mixed effects models. See Diggle, et al (2002), Verbeke and Molenberghs (2000, 2005) for an overview.

In many situations, however, we are interested in testing whether or not some of the between-subject variation is absent in a mixed effects model. This is equivalent to testing some variance components equal to zero. However, such a null hypothesis places some variance components on the boundary of the parameter space. Hence the commonly used tests, such as the likelihood ratio, Wald and score tests, do not have the traditional chi-squared distribution. In this chapter, we will review the likelihood ratio test and the score test for testing variance components in generalized linear mixed models.

A closely related topic is testing whether or not a covariate effect in a GLMM can be adequately represented by a polynomial of certain degree. Using a smoothing spline or penalized spline approach, testing for a polynomial covariate effect is equivalent to testing a zero variance component in an induced GLMM. We will review the likelihood ratio test and the score test for testing a parametric polynomial model versus a smoothing spline model for longitudinal data within the generalized additive mixed models framework (Lin and Zhang, 1999).

This chapter is organized as follows. In Section 2, we present the model specification of a GLMM and briefly review model estimation and inference procedures. In Section 3,

we review the likelihood ratio test for variance components in GLMMs and illustrate such tests in several common cases of interest. In Section 4, we review the score test for variance components in GLMMs, and compare the performance of the likelihood ratio test with the score test in a simple GLMM. In Section 6, we review the likelihood ratio test and the score test for testing a polynomial covariate effect versus a nonparametric smoothing spline model for longitudinal data. We illustrate these tests in Section 7 through application to data from Indonesian children infectious disease study. The chapter ends in Section 8 with a discussion.

2 Generalized Linear Mixed Models for Longitudinal/Clustered Data

Suppose there are m subjects in the sample. For the i th subject, denote by y_{ij} the response measured for the j th observation, e.g., the j th time point for longitudinal data or the j th outcome for multiple outcome data. Similarly denote by x_{ij} a $p \times 1$ vector of covariates associated with fixed effects and by z_{ij} a $q \times 1$ vector of covariate values associated with random effects. Given subject-specific random effects b_i , the responses y_{ij} are assumed to be conditionally independent and belong to an exponential family with the conditional mean $E(y_{ij}|b_i) = \mu_{ij}$ and conditional variance $\text{var}(y_{ij}|b_i) = V(\mu_{ij}) = \phi\omega_{ij}^{-1}v(\mu_{ij})$, where ϕ is a positive dispersion parameter, ω_{ij} is a pre-specified weight such as the binomial denominator when y_{ij} is the proportion of events in binomial sampling, and $v(\cdot)$ is the variance function. A generalized linear mixed model (GLMM) relates the conditional mean μ_{ij} to the covariates x_{ij} and z_{ij} as follows

$$g(\mu_{ij}) = x_{ij}^T\beta + z_{ij}^Tb_i, \tag{1}$$

where $g(\cdot)$ is a strictly increasing link function, β is a $p \times 1$ vector of fixed effects (regression coefficients) of x , and b_i is a $q \times 1$ vector of subject-specific random effects of z . The model specification is completed by the usual assumption that $b_i \sim N\{0, D(\psi)\}$, where ψ is a $c \times 1$ vector of variance components.

Model (1) includes many popular models for continuous and discrete data as special cases. For example, if the y_{ij} are continuous outcome measurements assumed to have a normal distribution given random effects b_i and the link function is the identity link $g(\mu) = \mu$, then model (1) reduces to the following linear mixed model (Laird and Ware, 1982)

$$y_{ij} = x_{ij}^T \beta + z_{ij}^T b_i + \epsilon_{ij}, \quad (2)$$

where $\epsilon_{ij} \stackrel{iid}{\sim} N(0, \phi)$ are residual errors. When the y_{ij} are binary responses, a common choice of the link function is the logit link $g(\mu) = \log\{\mu/(1 - \mu)\}$. In this case, model (1) reduces to the following logistic-normal model

$$\text{logit}\{P(y_{ij} = 1|b_i)\} = x_{ij}^T \beta + z_{ij}^T b_i. \quad (3)$$

The log-likelihood function $\ell(\beta, \psi; y)$ given outcome y under Model (1) is

$$\exp\{\ell(\beta, \psi; y)\} \propto |D(\psi)|^{-m/2} \prod_{i=1}^m \int \exp\left\{\sum_{j=1}^{n_i} \ell_{ij}(\beta, \psi; y_{ij}|b_i) - \frac{1}{2} b_i^T D^{-1}(\psi) b_i\right\} db_i, \quad (4)$$

where

$$\ell_{ij}(\beta, \psi; y_{ij}|b_i) = \int_{y_{ij}}^{\mu_{ij}} \frac{\omega_{ij}(y_{ij} - u)}{\phi v(u)} du$$

is the conditional log-likelihood of y_{ij} given random effects b_i .

Estimation and inference in Model (1) are often hampered by the intractable integrations involved in evaluation of likelihood (4) and have been well developed in the past two decades. Our main focus in this paper is on variance component testing in a GLMM. We hence list here some representative work as references. Zeger and Karim (1991) used a Gibbs sampling approach for model estimation and inference. Breslow and Clayton (1993) approximated the likelihood (4) using Laplace approximation and conducted model estimation and inference by maximizing a penalized quasi-likelihood (PQL). Breslow and Lin (1995) and Lin and Breslow (1996) studied the bias in PQL estimators and developed bias-correction methods. Booth and Hobert (1999) proposed an automated Monte Carlo EM algorithm to maximize the integrated likelihood (4).

As usual, throughout this chapter, we will use X for the design matrix of β and Z the design matrix of b . That is, $X = (X_1^T, X_2^T, \dots, X_m^T)^T$ where $X_i = (x_{i1}, x_{i2}, \dots, x_{in_i})^T$, and $Z = \text{diag}\{Z_1, Z_2, \dots, Z_m\}$ where $Z_i = (z_{i1}, z_{i2}, \dots, z_{in_i})^T$.

3 The Likelihood Ratio Test for Variance Components in GLMMs

The specification of the subject-specific random effects b_i in model (1) models the source of between-subject variation in the covariate effects of z , which also determines the within-subject correlation. The magnitude of this between-subject variation/within-subject correlation is captured by the magnitude of the elements of $D(\psi)$. In practice, investigators may be interested to see if there is no between-subject variation in some covariate effects of z . Statistically, it is equivalent to testing some or all of the elements of $D(\psi)$ to be zero.

In a regular hypothesis testing setting, a likelihood ratio test (LRT) is the most commonly used test due to its desirable theoretical properties and the fact that it is easy to construct. Under very general regularity conditions, the LRT statistic asymptotically has a χ^2 null distribution with the degrees of freedom equal to the number of independent parameters being tested under the null hypothesis. However, when the elements of $D(\psi)$ are tested, the null hypothesis usually places some or all of the components of ψ on the boundary of the model parameter space, in which case the LRT statistic does not have the usual χ^2 null distribution.

Denote by $\theta = (\beta^T, \psi^T)^T$, a combined vector of regression and variance-covariance parameters in the model. Self and Liang (1987) formulated the asymptotic null distribution of the LRT statistic $-2\ln\lambda_m$ for testing

$$H_0 : \theta_0 \in \Omega_0 \text{ vs. } H_A : \theta_0 \in \Omega_1 = \Omega \setminus \Omega_0,$$

when the true value θ_0 of θ is possibly on the boundary of the model parameter space Ω . Assume that the parameter spaces Ω_1 under H_A and Ω_0 under H_0 can be approximated at

θ_0 by cones C_{Ω_1} and C_{Ω_0} , respectively, with vertex θ_0 . Self and Liang (1987) showed that under some regularity conditions the LRT statistic $-2\ln\lambda_m$ asymptotically has the same distribution as

$$\inf_{\psi \in C_{\Omega_0} - \theta_0} \{(U - \theta)^T I(\theta_0)(U - \theta)\} - \inf_{\psi \in C_{\Omega} - \theta_0} \{(U - \theta)^T I(\theta_0)(U - \theta)\}, \quad (5)$$

where C_{Ω} is the cone approximating Ω with vertex at θ_0 , $C_{\Omega} - \theta_0$ and $C_{\Omega_0} - \theta_0$ are translated cones of C_{Ω} and C_{Ω_0} such that their vertices are the origin, $I(\theta_0)$ is the (Fisher) information matrix at θ_0 , and U is a random vector distributed as $N\{0, I^{-1}(\theta_0)\}$. Alternatively, Self and Liang (1987) expressed (5) as

$$\inf_{\psi \in \tilde{C}_0} \|\tilde{U} - \theta\|^2 - \inf_{\psi \in \tilde{C}} \|\tilde{U} - \theta\|^2, \quad (6)$$

where $\tilde{C} = \{\tilde{\theta} : \tilde{\theta} = \Lambda^{1/2}Q^T\theta \text{ for all } \theta \in C_{\Omega} - \theta_0\}$, $\tilde{C}_0 = \{\tilde{\theta} : \tilde{\theta} = \Lambda^{1/2}Q^T\theta \text{ for all } \theta \in C_{\Omega_0} - \theta_0\}$, \tilde{U} is a random vector from $N(0, I)$ and $Q\Lambda Q^T$ is the spectral decomposition of $I(\theta_0)$; that is, $I(\theta_0) = Q\Lambda Q^T$, $QQ^T = I$ and $\Lambda = \text{diag}\{\lambda_i\}$. We can use either (5) or (6) to derive the asymptotic null distribution for the LRT statistic depending on the structure of $I(\theta_0)$.

Stram and Lee (1994) applied the above general results of Self and Liang (1987) to investigate the asymptotic null distribution of LRT statistic $-2\ln\lambda_m$ for testing components of $D(\theta)$ for linear mixed model (2). Since the results of Self and Liang (1987) are for a general parametric model, they are also applicable to GLMM (1) as long as one can maximize the likelihood (4) under the null and alternative hypotheses of interest. Here we list some cases one commonly encounters in practice. For reviews on LRT for variance components in linear mixed models, see the chapter ‘‘Likelihood ratio testing for zero variance components in linear mixed models’’ by Crainiceanu.

Case 1: Assume the dimension q of the random effects is equal to one, that is, $D = d_{11}$, and we are testing $H_0 : d_{11} = 0$ vs. $H_A : d_{11} > 0$. For example, consider the random intercept model $Z_{ij}b_i = b_i$ and $b_i \sim N(0, d_{11})$ in model (1).

In this case, $\theta = (\beta^T, d_{11})^T$ and $C_{\Omega_0} = R^p \times \{0\}$ and $C_{\Omega_1} = R^p \times (0, \infty)$. Decompose U and $I(\theta_0)$ in (5) as $U = (U_1^T, U_2)^T$ and $I(\theta_0) = \{I_{jk}\}$ corresponding to β and d_{11} . Some algebra then shows that

$$\inf_{\psi \in C_{\Omega_0} - \theta_0} \{(U - \theta)^T I(\theta_0)(U - \theta)\} = \tilde{U}_2^2,$$

where $\tilde{U}_2 = (I_{22} - I_{21}I_{11}^{-1}I_{12})^{1/2}U_2$, and

$$\inf_{\psi \in C_{\Omega} - \theta_0} \{(U - \theta)^T I(\theta_0)(U - \theta)\} = \tilde{U}_2^2 I(\tilde{U}_2 \leq 0).$$

Therefore, (5) reduces to $\tilde{U}_2^2 I(\tilde{U}_2 > 0)$. It is easy to see that $\tilde{U}_2 \sim N(0, 1)$. The asymptotic null distribution of $-2\ln\lambda_m$ (as $m \rightarrow \infty$) is then a 50:50 mixture of χ_0^2 and χ_1^2 .

Denote the observed LRT statistic by T_{obs} . Then the level α likelihood ratio test will reject $H_0 : d_{11} = 0$ if $T_{obs} \geq \chi_{2\alpha, 1}^2$, where $\chi_{2\alpha, 1}^2$ is the $(1 - 2\alpha)$ th quantile of the χ^2 distribution with one degree of freedom. The corresponding p-value is $P[\chi_1^2 \geq T_{obs}]/2$, half of the p-value if the regular but incorrect χ_1^2 distribution were used.

Case 2: Assume $q = 2$ so $D = \{d_{ij}\}_{2 \times 2}$, and we test $H_0 : d_{11} > 0, d_{12} = d_{22} = 0$ vs. $H_A : D$ is positive definite. As an example, consider the random intercept and slope model $z_{ij}^T b_i = b_{0i} + b_{1i}t_{ij}$, where t_{ij} is time and b_{0i} and b_{1i} are subject-specific random intercept and slope in longitudinal data assumed to follow $(b_{0i}, b_{1i}) \sim N\{0, D(\psi)\}$. The foregoing hypothesis tests the random intercept model (H_0) versus the random intercept and slope model (H_1).

In this case, $\theta = (\theta_1^T, \theta_2, \theta_3)^T$ where $\theta_1 = (\beta^T, d_{11})^T$, $\theta_2 = d_{12}$ and $\theta_3 = d_{22}$. Under $H_0 : d_{11} > 0$, the translated approximating cone at θ_0 is $C_{\Omega_0} - \theta_0 = R^{p+1} \times \{0\} \times \{0\}$. Under $H_0 \cup H_A$, $d_{11} > 0$ and D is positive semidefinite. This is equivalent to $d_{11} > 0$ and $d_{22} - d_{11}^{-1}d_{12}^2 \geq 0$. Since the boundary defined by $d_{22} - d_{11}^{-1}d_{12}^2 = 0$ for any given $d_{11} > 0$ is a smooth surface, the translated approximating cone at θ_0 under $H_0 \cup H_A$ is $C_{\Omega} - \theta_0 = R^{p+1} \times R^1 \times [0, \infty)$. Similar to *Case 1*, decompose U and $I^{-1}(\theta_0)$ in (5) as $U = (U_1^T, U_2, U_3)^T$ and $I^{-1}(\theta_0) = \{I^{jk}\}$ corresponding to θ_1 , θ_2 and θ_3 . We can then show

that

$$\inf_{\psi \in C_{\Omega_0} - \theta_0} \{(U - \theta)^T I(\theta_0)(U - \theta)\} = [U_2, U_3] \begin{bmatrix} I^{22} & I^{23} \\ I^{32} & I^{33} \end{bmatrix}^{-1} \begin{bmatrix} U_2 \\ U_3 \end{bmatrix}, \quad (7)$$

$$\inf_{\psi \in C_{\Omega} - \theta_0} \{(U - \theta)^T I(\theta_0)(U - \theta)\} = (I^{33})^{-1} U_3^2 I(U_3 \leq 0). \quad (8)$$

Since $(U_1^T, U_2, U_3)^T \sim N\{0, I^{-1}(\theta_0)\}$, the distribution of the difference between (7) and (8) is a 50:50 mixture of χ_1^2 and χ_2^2 .

For a given significance level α , the critical value c_α for the LRT can be solved by the following equation using some statistical software

$$0.5P[\chi_1^2 \geq c] + 0.5P[\chi_2^2 \geq c] = \alpha.$$

Alternatively, the significance level α can also be compared to the LRT p-value

$$\text{p-value} = 0.5P[\chi_1^2 \geq T_{obs}] + 0.5P[\chi_2^2 \geq T_{obs}],$$

where T_{obs} is the observed LRT statistic. This p-value is always smaller than the usual but incorrect p-value $P[\chi_2^2 \geq T_{obs}]$ in this setting. The decision based on this classical p-value is hence conservative.

Case 3: Assume $q > 2$ and we test the presence of the q th element of the random effects b_i in model (1). Denote $D = \begin{pmatrix} D_{11} & D_{12} \\ D_{21} & D_{22} \end{pmatrix}$, where the dimensions of D_{11} , D_{12} and D_{21} are $s \times s$, $s \times 1$ and $1 \times s$ respectively ($s = q - 1$), and D_{22} is a scalar. Then statistically, we test $H_0 : D_{11}$ is positive definite, $D_{12} = 0, D_{22} = 0$ vs. $H_A : D$ is positive definite.

Denote by θ_1 the combined vector of β and the unique elements of D_{11} , $\theta_2 = D_{12}$ and $\theta_3 = D_{22}$. Under H_0 , the translated approximating cone at θ_0 is $C_{\Omega_0} - \theta_0 = R^{p+s(s+1)/2} \times \{0\}^s \times \{0\}$. Under $H_0 \cup H_A$, D_{11} is positive definite and D is positive semidefinite. This is equivalent to D_{11} being positive definite and $D_{22} - D_{12}^T D_{11}^{-1} D_{12} \geq 0$ (Stram and Lee, 1994, incorrectly used q constraints). Again, since the boundary defined by $D_{22} - D_{12}^T D_{11}^{-1} D_{12} = 0$ for any given positive definite matrix D_{11} is a smooth surface, the translated approximating cone at θ_0 under $H_0 \cup H_A$ is $C_{\Omega} - \theta_0 = R^{p+s(s+1)/2} \times R^s \times [0, \infty)$. This case is similar to *Case 2*

except that U_2 is an $s \times 1$ random vector. Therefore, the asymptotic null distribution of LRT statistic is a 50:50 mixture of χ_s^2 and χ_{s+1}^2 . The p-value of the LRT test for given observed LRT statistic T_{obs} is equal to $0.5P[\chi_s^2 \geq T_{obs}] + 0.5P[\chi_{s+1}^2 \geq T_{obs}]$, which will be closer to the usual but incorrect p-value $P[\chi_{s+1}^2 \geq T_{obs}]$ as s becomes larger.

Case 4: Suppose the random effects part $z_{ij}^T b_i$ in model (1) can be decomposed as $z_{ij}^T b_i = z_{1ij}^T b_{1i} + z_{2ij}^T b_{2i}$, where $b_{1i} \sim N\{0, D_1(\psi_1)\}$, $b_{2i} \sim N(0, \psi_2 I)$ and we test $H_0 : \psi_2 = 0$ and D_1 is positive definite versus $H_A : \psi_2 > 0$ and D_1 is positive definite. Denote by θ_1 the combined vector of β and the unique elements of D_1 , and $\theta_2 = \psi_2$. Since the true values of the nuisance parameters θ_1 are interior points of the corresponding parameter space, we can apply the result of *Case 1* to this case. This implies that the asymptotic null distribution of the LRT statistic is a 50:50 mixture of χ_0^2 and χ_1^2 .

Case 5: Suppose $D_1(\psi_1)$ in *Case 4* takes the form $\psi_1 I$, and we test $H_0 : \psi_1 = 0, \psi_2 = 0$ versus $H_A : \text{either } \psi_1 > 0 \text{ or } \psi_2 > 0$. Denote $\theta = (\beta^T, \psi_1, \psi_2)$ with $\theta_1 = \beta$, $\theta_2 = \psi_1$ and $\theta_3 = \psi_2$. Under H_0 , the translated approximating cone at θ_0 is $C_{\Omega_0} - \theta_0 = R^p \times \{0\} \times \{0\}$. Under $H_0 \cup H_A$, the translated approximating cone at θ_0 is $C_{\Omega} - \theta_0 = R^p \times [0, \infty) \times [0, \infty)$.

Decompose U and $I(\theta_0)$ in (5) as $(U_1^T, U_2, U_3)^T$ and $I(\theta_0) = \{I_{ij}\}$ corresponding to θ_1, θ_2 and θ_3 , and define matrix \tilde{I} as follows:

$$\tilde{I} = \begin{bmatrix} \tilde{I}_{22} & \tilde{I}_{23} \\ \tilde{I}_{32} & \tilde{I}_{33} \end{bmatrix} = \begin{bmatrix} I_{22} & I_{23} \\ I_{32} & I_{33} \end{bmatrix} - \begin{bmatrix} I_{21} \\ I_{31} \end{bmatrix} I_{11}^{-1} [I_{12}, I_{13}].$$

Then $(U_2, U_3)^T \sim N(0, \tilde{I}^{-1})$. Given θ_2 and θ_3 , it can be easily shown that

$$\inf_{\theta_1 \in R^p} (U - \theta)^T I(\theta_0) (U - \theta) = [U_2 - \theta_2, U_3 - \theta_3] \tilde{I} \begin{bmatrix} U_2 - \theta_2 \\ U_3 - \theta_3 \end{bmatrix} = (\tilde{U}_2 - \tilde{\theta}_2)^2 + (\tilde{U}_3 - \tilde{\theta}_3)^2,$$

where $(\tilde{U}_2, \tilde{U}_3)^T = \tilde{\Lambda}^{1/2} \tilde{Q}^T (Z_2, Z_3)^T$, $(\tilde{\theta}_2, \tilde{\theta}_3)^T = \tilde{\Lambda}^{1/2} \tilde{Q}^T (\theta_2, \theta_3)^T$, $\tilde{Q} \tilde{\Lambda} \tilde{Q}^T$ is the spectral decomposition of \tilde{I} . Therefore, under H_0 , we have

$$\inf_{\theta \in C_{\Omega_0} - \theta_0} (U - \theta)^T I(\theta_0) (U - \theta) = \tilde{U}_2^2 + \tilde{U}_3^2.$$

Denote by φ the angle in the radiant formed by the vectors $\tilde{\Lambda}^{1/2} \tilde{Q}(1, 0)^T$ and $\tilde{\Lambda}^{1/2} \tilde{Q}(0, 1)^T$, that is, $\varphi = \cos^{-1} \left(\tilde{I}_{23} / \sqrt{\tilde{I}_{22} \tilde{I}_{33}} \right)$ (Self and Liang, 1987, who incorrectly used I_{jk}), and set

$\xi = \varphi/2\pi$, then

$$\inf_{\theta \in C_{\Omega} - \theta_0} (U - \theta)^T I(\theta_0) (U - \theta) = \begin{cases} \tilde{U}_2^2 + \tilde{U}_3^2 & \text{with probability } \xi \\ \tilde{U}_2^2 & \text{with probability } 0.25 \\ \tilde{U}_3^2 & \text{with probability } 0.25 \\ 0 & \text{with probability } 0.5 - \xi. \end{cases}$$

Therefore, the asymptotic null distribution of the LRT statistic is a mixture of χ_0^2 , χ_1^2 and χ_2^2 with mixing probabilities ξ , 0.5 and $0.5 - \xi$. Note that since \tilde{I} is a positive definite matrix, the probability ξ satisfies $0 < \xi < 0.5$. In particular, if \tilde{I} is diagonal, the mixing probabilities are 0.25, 0.5 and 0.25.

The asymptotic null distribution of the LRT statistic is relatively easier to study for the above cases. The structure of the information matrix $I(\theta_0)$ and the approximating cones $C_{\Omega} - \theta_0$ and $C_{\Omega_0} - \theta_0$ play key roles in deriving the asymptotic null distribution. For more complicated cases of testing variance components, although the asymptotic null distribution of the LRT is generally still a mixture of some chi-squared distributions, it may be too difficult to derive the mixing probabilities. In this case one may use simulation to calculate the p-value.

4 The Score Test for Variance Components in GLMMs

Conceptually, the LRT test for variance components in GLMMs discussed in Section 3 is easy to apply. However, the LRT involves fitting GLMM (1) under H_0 and $H_0 \cup H_A$. For many situations, it is relatively straightforward to fit model (1) under H_0 . However, one could often encounter numerical difficulties in fitting the full model (1) under $H_0 \cup H_A$. First, fitting model (1) under $H_0 \cup H_A$ involves higher dimensional integration, thus increasing computational burden. Second, if H_0 is true or approximately true, it is often unstable to fit a more complicated model under $H_0 \cup H_A$ as the parameters used to specify H_0 are estimated close to the boundary. For example, although the Laplace approximation used by Breslow and Clayton (1993) and others is recommended for a GLMM with complex

parameter boundary, such approximation may work poorly in such cases (Hsiao, 1997). In this section, we discuss score tests for variance components in model (1). One advantage of using score tests is that we only need to fit model (1) under H_0 , often dramatically reducing computational burden. Another advantage is that unlike likelihood ratio tests, score tests only require the specification of the first two moments of random effects and are hence robust to mis-specification of the distribution of random effects (Lin, 1997).

We first review the score test for *Case 1* discussed in Section 3, that is, we assume there is only one variance component in model (1) for which we would like to conduct hypothesis testing. A one-sided score test is desirable in this case and can be found in Lin (1997), Jacqmin-Gadda and Commenges (1995). Zhang (1997) discussed a one-sided score test for testing $H_0 : \psi_2 = 0$ for *Case 5* in Section 3 for a generalized additive mixed model, which includes model (1) as a special case. Verbeke and Molenberghs (2003) discussed one-sided score tests for linear mixed model (2). Lin (1997) derived score statistics for testing single or multiple variance components in GLMMs and considered simpler two-sided tests. Parallel to likelihood ratio tests, the one-sided score tests follow a mixture of chi-square distribution whose weights could be difficult to calculate when multiple variance components are set to be zero under H_0 as illustrated in *Case 5*. The two-sided score tests assume the score statistic follows a regular chi-square distribution and hence its p-value can be calculated more easily, especially for multiple variance component tests. The two-sided score test has the correct size under H_0 , while its power might be lower than the one-sided score and likelihood ratio tests. See the simulation results for more details.

In *Case 1*, $\psi = d_{11}$. Assume at the moment that β is known. One can show using L'Hôpital's rule or the Taylor expansion (Lin, 1997) that the score for ψ is

$$U_\psi = \frac{\partial \ell(\beta, \psi; y)}{\partial \psi} \Big|_{\psi=0} = \frac{1}{2} \sum_{i=1}^m \left[\left\{ \sum_{j=1}^{n_i} z_{ij} w_{ij} \delta_{ij} (y_{ij} - \mu_{ij}^0) \right\}^2 - \sum_{j=1}^{n_i} z_{ij}^2 \{w_{ij} + e_{ij} (y_{ij} - \mu_{ij}^0)\} \right], \quad (9)$$

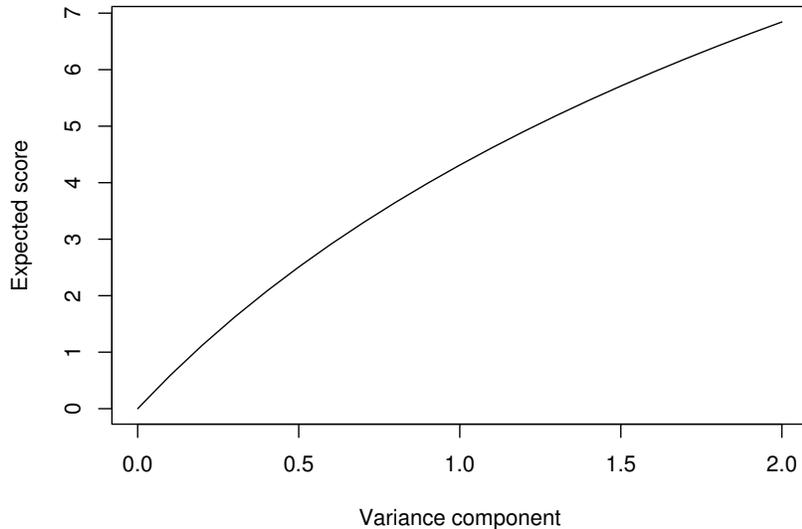
where $w_{ij} = [V(\mu_{ij}^0)\{g'(\mu_{ij}^0)\}^2]^{-1}$, $\delta_{ij} = g'(\mu_{ij}^0)$,

$$e_{ij} = \frac{V'(\mu_{ij}^0)g'(\mu_{ij}^0) + V(\mu_{ij}^0)g''(\mu_{ij}^0)}{V^2(\mu_{ij}^0)\{g'(\mu_{ij}^0)\}^3},$$

which is zero for the canonical link function $g(\cdot)$, and μ_{ij}^0 satisfies $g(\mu_{ij}^0) = x_{ij}^T\beta$.

It can be easily shown that the random variable U_ψ defined by (9) has zero mean under $H_0 : \psi = 0$. As argued by Verbeke and Molenberghs (2003), the log-likelihood $\ell(\beta, \psi; y)$ for the linear mixed model (2) on average has a positive slope at $\psi = 0$ when in fact $\psi > 0$. The same argument also applies to GLMM (1). This is because under $H_A : \psi > 0$, the MLE $\hat{\psi}$ of ψ will be close to ψ so that $\hat{\psi} > 0$ when the sample size m gets large. If the log-likelihood $\ell(\beta, \psi; y)$ as a function of ψ only is smooth and has a unique MLE $\hat{\psi}$, which is the case for most GLMMs, the slope U_ψ of $\ell(\beta, \psi; y)$ at $\psi = 0$ will be positive. Indeed, $E(U_\psi)$ generally is an increasing function of ψ . For example, Figure 1 plots the expected score $E(U_\psi)$ vs. ψ for the logistic-normal GLMM (3) where $m = 10$, $n_i = 5$, $x_{ij} = 1$, $\beta = 0.25$ and $z_{ij} = 1$. It is confirmed that $E(U_\psi)$ increases as ψ increases.

Figure 1: *Expected score as a function of variance component ψ*



The above argument indicates that a large value of U_ψ provides evidence against $H_0 :$

$\psi = 0$ and we should reject H_0 only if U_ψ is large. Since U_ψ is a sum of independent random variables, classic results show that it will have an asymptotic normal distribution under $H_0 : \psi = 0$ with zero mean and variance equal to $I_{\psi\psi} = E(U_\psi^2)$, where the expectation is taken at $H_0 : \psi = 0$.

Denote by κ_{rij} the r th cumulant of y_{ij} under H_0 . By the properties of the distributions in an exponential family, κ_{3ij} and κ_{4ij} are related to κ_{2ij} via $\kappa_{(r+1)ij} = \kappa_{2ij} \partial \kappa_{rij} / \partial \mu_{ij}$ ($r = 2, 3$), where $\kappa_{2ij} = \phi \omega_{ij}^{-1} v(\mu_{ij})$ and $\mu_{ij} = \mu_{ij}^0$. Specifically,

$$\kappa_{3ij} = (\phi \omega_{ij}^{-1})^2 v'(\mu_{ij}) v(\mu_{ij}), \quad \kappa_{4ij} = (\phi \omega_{ij}^{-1})^3 [v''(\mu_{ij}) v(\mu_{ij}) + \{v'(\mu_{ij})\}^2] v(\mu_{ij}).$$

Then $I_{\psi\psi}$ can be shown to be (Lin, 1997)

$$I_{\psi\psi} = \frac{1}{4} \sum_{i=1}^m \sum_{j=1}^{n_i} z_{ij}^2 r_{ii},$$

where $r_{ii} = w_{ij}^4 \delta_{ij}^4 \kappa_{4ij} + 2w_{ij}^2 + e_{ij} \kappa_{2ij} - 2w_{ij}^2 \delta_{ij}^2 e_{ij} \kappa_{3ij}$. Therefore, a level α score test for testing $H_0 : \psi = 0$ vs. $H_A : \psi > 0$ will reject $H_0 : \psi = 0$ if $U_\psi \geq z_\alpha I_{\psi\psi}^{1/2}$.

In practice, however, β in U_ψ and $I_{\psi\psi}$ is unknown and has to be estimated under H_0 . This is straightforward since under $H_0 : \psi = 0$, GLMM (1) reduces to the standard generalized linear model for independent data $g(\mu_{ij}) = X_{ij}^T \beta$ and existing software can be used to easily calculate the MLE $\hat{\beta}$ of β under $H_0 : \psi = 0$. In this case, Lin (1997) considered the bias-corrected score statistic to account for the estimation of β under H_0 as

$$U_\psi^c = \left. \frac{\partial \ell(\beta, \psi; \mathbf{y})}{\partial \psi} \right|_{\psi=0, \beta=\hat{\beta}} = \frac{1}{2} \sum_{i=1}^m \left[\left\{ \sum_{j=1}^{n_i} z_{ij} w_{ij} \delta_{ij} (y_{ij} - \hat{\mu}_{ij}^0) \right\}^2 - \sum_{j=1}^{n_i} z_{ij}^2 w_{0ij} \right], \quad (10)$$

where all quantities are obtained by replacing β by $\hat{\beta}$, $w_{0ij} = (1 - h_{ij})w_{ij} + e_{ij}(y_{ij} - \hat{\mu}_{ij}^0)$, and h_{ij} is the corresponding diagonal element of the hat matrix $H = W^{1/2} X (X W X)^{-1} X^T W^{1/2}$, $W = \text{diag}\{w_{ij}\}$, and showed that U_ψ^c has variance

$$\tilde{I}_{\psi\psi} = I_{\psi\psi} - I_{\psi\beta}^T I_{\beta\beta}^{-1} I_{\psi\beta}, \quad (11)$$

where

$$I_{\psi\beta} = \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^{n_i} c_{ij} z_{ij} x_{ij}, \quad I_{\beta\beta} = X^T W X = \sum_{i=1}^m \sum_{j=1}^{n_i} w_{ij} x_{ij} x_{ij}^T \quad (12)$$

with $c_{ij} = w_{ij}^3 \delta_{ij}^3 \kappa_{3ij} - w_{ij} \delta_{ij} e_{ij} \kappa_{2ij}$. Then the bias-corrected score test at level α would reject H_0 if $T_s = U_\psi^c \geq z_\alpha \tilde{I}_{\psi\psi}^{1/2}$. The one-sided score test presented above is asymptotically equivalent to the likelihood ratio test (Verbeke and Molenberghs, 2003). The two-sided score test assumes the score statistic $T_s = \{U_\psi^c\}^2 / \tilde{I}_{\psi\psi}$ follows a χ^2 distribution. Unlike the regular likelihood ratio test, such a two-sided score test has the correct size under H_0 but is subject to some loss of power. As shown in our simulation studies for a single variance component, the loss of power is minor to moderate for most alternatives. The highest power loss is about 10% when the magnitude of the variance component is moderate.

When the dimension of ψ is greater than 1, suppose we can partition $\psi = (\psi_1, \psi_2)$ where ψ_1 is a $c_1 \times 1$ vector and ψ_2 is a $c_2 \times 1$ vector. We are interested in testing $H_0 : \psi_1 = 0$ vs. $H_A : \psi_1 \geq 0$. Here the inequality is interpreted element-wise. Lin (1997) considered a simple two-sided score test for this multiple variance component test. Specifically, denote by $(\hat{\beta}, \hat{\psi}_2)$ the MLE of (β, ψ_2) under $H_0 : \psi_1 = 0$. We can similarly derive the (corrected) score $S_{\psi_1} = m^{-1/2} \partial \ell(\beta, \psi_1; y) / \partial \psi_1 |_{\psi_1=0, \beta=\hat{\beta}, \psi_2=\hat{\psi}_2}$. See Lin (1997) for the special case where each element of ψ represents a variance of a random effect. Asymptotically, S_ψ has a normal distribution with zero mean and variance equal to the efficient information matrix $H_{\psi_1\psi_1} = m^{-1} \tilde{I}_{\psi_1\psi_1}$ under H_0 , where $\tilde{I}_{\psi_1\psi_1}$ is defined similarly to (11) except that $I_{\phi\beta}$ and $I_{\beta\beta}$ are replaced by $I_{\psi_1\gamma}$ and $I_{\gamma\gamma}$ and $\gamma = (\psi_2, \beta)$. The simple two-sided score statistic is defined as

$$T_s = S_{\psi_1}^T H_{\psi_1\psi_1}^{-1} S_{\psi_1}, \quad (13)$$

and the p-value is calculated by assuming T_s follows a chi-square distribution with c_1 degrees of freedom.

Silvapulle and Silvapulle (1995) proposed a one-sided score test for a general parametric model and showed that the one-sided score test is asymptotically equivalent to the likelihood

ratio test. Verbeke and Molenberghs (2003) extended Silvapulle and Silvapulle’s (1995) one-sided score test for testing variance components $H_0 : \psi_1 = 0$ vs. $H_A : \psi_1 \in \mathcal{C}$ for linear mixed model (2) and showed similar asymptotic equivalence between the one-sided score test and the likelihood ratio test. Hall and Praestgaard (2001) derived a one-sided score test for GLMMs. Then the one-sided score statistic T_s^* is defined as

$$T_s^* = S_{\psi_1}^T H_{\psi_1 \psi_1}^{-1} S_{\psi_1} - \inf_{\psi_1 \in \mathcal{C}} \{(S_{\psi_1} - \psi_1)^T H_{\psi_1 \psi_1}^{-1} (S_{\psi_1} - \psi_1)\}. \quad (14)$$

It is easy to see that T_s^* as defined in (14) has the same asymptotic null distribution of the likelihood ratio test for testing $H_0 : \psi_1 = 0$ vs. $H_A : \psi_1 \in \mathcal{C}$. Similarly to the case for the likelihood ratio test, it is critical to determine $H_{\psi\psi}$ and the shape of \mathcal{C} , and T_s^* generally follows a mixture of chi-square distributions and we usually have to study the distribution of T_s^* case by case. Both the two-sided test T_s and the one-sided test T_s^* have the correct size under H_0 . The two-sided test T_s is much easier to calculate, but is subject to some loss of power. Hall and Praestgaard (2001) conducted extensive simulation studies comparing Lin’s (1997) two-sided score test and their one-sided score test for GLMMs with two-dimensional random effects and found similar power loss to the case of a single variance component (Table 4 in Hall and Praestgaard, 2001; the maximum power loss is about 9%).

5 Simulation Study to Compare the Likelihood Ratio Test and the Score Test for Variance Components

We conducted a small simulation study to compare the size and the power of the one-sided and two-sided score tests with the likelihood ratio test. We considered the logistic-normal GLMM (3) by assuming binary responses y_{ij} ($i = 1, 2, \dots, m = 100$, $j = 1, 2, \dots, n_i = 5$) were generated from the following logistic-normal GLMM

$$\text{logitP}(y_{ij} = 1 | b_i) = \beta + b_i, \quad (15)$$

where $\beta = 0.25$ and $b_i \sim N(0, \psi)$, with equal spaced ψ in $[0,1]$ by 0.2. For each value of ψ , 500 data sets were generated. The likelihood ratio test described in Section 3 and the (corrected) one-sided and two-sided score tests were applied to test $H_0 : \psi = 0$. We compare the performance of the regular but conservative LRT, the appropriate LRT, one-sided and two-sided score test for testing $H_0 : \psi = 0$. The nominal level of all 4 tests were set at $\alpha = 0.05$.

Table 1: *Size and Power comparisons of the likelihood ratio tests and score tests for a single variance component based on 500 simulations under the logistic model (15)*

Method	Size		Power			
	$\psi = 0$	$\psi = 0.2$	$\psi = 0.4$	$\psi = 0.6$	$\psi = 0.8$	$\psi = 1.0$
LRT	0.034	0.370	0.790	0.922	0.990	1.000
Regular LRT	0.020	0.280	0.672	0.882	0.968	0.992
One-sided score test	0.054	0.416	0.834	0.938	0.996	1.000
Two-sided score test	0.050	0.336	0.736	0.910	0.980	0.998

Table 5 presents the simulation results. The results show that the size of the (correct) likelihood ratio test is little smaller than the nominal level. This is probably due to the numerical instability caused by numerical difficulties in fitting model (15) when in fact there is no random effect in the model, or the fact that the sample size (number of clusters $m = 100$) may not be large enough for the asymptotic theory to take effect. As expected, the regular LRT using χ_1^2 is conservative and the size is too small. On the other hand, both one-sided and two-sided score test have their sizes very close to the nominal level. The powers of the likelihood ratio test and the one-sided score test are almost the same, although the one-sided score test is slightly more powerful than the LRT, which may be due to the numerical integration required to fit model (15). The two-sided score test has some loss of power compared to the one-sided score test and the correct LRT. However the p-value of the two-sided score test is much easier to calculate especially for testing for multiple variance components.

6 Polynomial Test in Semiparametric Additive Mixed Models

Lin and Zhang (1999) proposed generalized additive mixed models (GAMMs), an extension of GLMMs where each parametric covariate effect in model (1) is replaced by a smooth but arbitrary nonparametric function, and proposed to estimate each function by a smoothing spline. Using a mixed model representation for a smoothing spline, they cast estimation and inference of GAMMs in a unified framework through a working GLMM, where the inverse of a smoothing parameter is treated as a variance component. A special case of GAMMs is the semiparametric additive mixed models considered by Zhang and Lin (2003):

$$g(\mu_{ij}) = f(t_{ij}) + s_{ij}^T \alpha + z_{ij}^T b_i, \quad (16)$$

where $f(t)$ is an unknown smooth function, i.e., the covariate effect of t is assumed to be nonparametric, s_{ij} is some covariate vector, and $b_i \sim N\{0, D(\psi)\}$. For independent (normal) data with the identity link, model (16) reduces to a partially linear model. We are interested in developing a score testing for testing $f(t)$ is a parametric polynomial function versus a smooth nonparametric function. Specifically, we set H_0 : $f(t)$ is a polynomial function of degree $K - 1$ and H_1 : $f(t)$ is a smoothing spline.

Following Zhang and Lin (2003), denote by $t^0 = (t_1^0, t_2^0, \dots, t_r^0)^T$ a vector of ordered distinct t_{ij} 's and by f a vector of $f(t)$ evaluated at t^0 (without loss of generality, assume $0 < t_1^0 < \dots < t_r^0 < 1$). The K th-order ($K \geq 1$) smoothing spline estimator $f(t)$ can be expressed as

$$f(t) = \sum_{k=1}^K \delta_k \phi_k(t) + \sum_{l=1}^r a_l R(t, t_l^0), \quad (17)$$

where $\{\phi_k(t) = t^{k-1}/(k-1)!\}_{k=1}^K$ is a basis for the space of polynomials of order $K - 1$ and $R(t, s)$ is defined as

$$R(t, s) = \frac{1}{\{(K-1)!\}^2} \int_0^1 (s-u)_+^{K-1} (t-u)_+^{K-1} du.$$

Then the smoothing spline estimator of f has the following mixed effect representation

$$f = T\delta + \Sigma a, \quad (18)$$

where T is an $r \times K$ matrix with the (l, k) th element equal to $\phi_k(t_l^0)$, Σ is a positive matrix with the (l, k) th element equal to $R(t_l^0, t_k^0)$, $\delta = (\delta_1, \dots, \delta_K)^T$ is a vector of fixed effects and $a = (a_1, a_2, \dots, a_r)^T \sim N(0, \tau\Sigma^{-1})$ is a vector of random effects with $\tau \geq 0$ being the inverse of the smoothing parameter for the smoothing spline estimate $f(t)$.

Let $n = \sum_{i=1}^m n_i$ be the total sample size and denote by N the $n \times r$ incidence matrix that maps $\{t_{ij}\}$'s into t^0 . Further denote $S = \{(s_{i1}, \dots, s_{in_i})^T\}$, $X = (NT, S)$, $B = N\Sigma$, $\mu^b = (\mu_{11}^b, \dots, \mu_{1n_1}^b, \dots, \mu_{m1}^b, \dots, \mu_{mn_m}^b)^T$. Then under the mixed effect representation (18), semiparametric additive mixed model (16) reduces to a generalized linear mixed model in matrix notation

$$g(\mu) = X\beta + Ba + Zb, \quad (19)$$

where $\beta = (\delta^T, \alpha^T)^T$ are new fixed effects and a and $b = (b_1^T, \dots, b_m^T)^T$ are independent new random effects. Therefore, the smoothing spline estimator $f(t)$ can be estimated using the estimation procedure for a GLMM, such as maximum penalized quasi-likelihood procedure of Breslow and Clayton (1993).

We are interested in using this spline and mixed model connection to test whether or not the smoothing spline $f(t)$ in semiparametric additive mixed model (16) can be adequately modeled by a polynomial of order $K - 1$, i.e., $H_0: f(t)$ is a polynomial of order $K - 1$ and $H_A: f(t)$ is a smoothing spline. From the smoothing spline expression (17), it is clear that $f(t)$ is a polynomial of order $K - 1$ if and only if $a_1 = a_2 = \dots = a_r = 0$. By mixed effect representation (18), this test is equivalent to the variance component test $H_0: \tau = 0$ vs $H_A: \tau > 0$. It is hence natural to consider using the variance component likelihood ratio test or score test described in the previous sections to test $H_0: \tau = 0$. However, the data do not have independent cluster structure under the alternative $H_A: \tau > 0$. Therefore,

the asymptotic null distribution of the likelihood ratio test statistic for testing $H_0 : \tau = 0$ does not follow a 50:50 mixture of χ_0^2 and χ_1^2 . In fact, for independent normal data with the identity link, Crainiceanu *et al.* (2005) showed that, when $f(t)$ is modeled by a penalized spline (similar to a smoothing spline), the LRT statistic asymptotically has approximately 0.95 mass probability at zero. For this special case, Crainiceanu *et al.* (2005) derived the exact null distribution of the LRT statistic. Their results, however, may not be applicable to testing $H_0 : \tau = 0$ under a more general mixed model representation (19). Furthermore, it could be computationally difficult to calculate this LRT statistic by fitting model (19) under the alternative $H_A : \tau > 0$ as it usually requires high-dimensional numerical integrations.

Due to the special structure of the smoothing matrix Σ , the score statistic of τ evaluated under $H_0 : \tau = 0$ does not have a normal distribution. Zhang and Lin (2003) showed that the score statistic of τ can usually be expressed as a weighted sum of chi-squared random variables with positive but rapidly decaying weights, and its distribution can be adequately approximated by that of a scaled chi-squared random variable.

Under the mixed model representation (19), the marginal likelihood function $L_M(\tau, \psi; y)$ of (τ, ψ) is given by

$$L_M(\tau, \psi; y) \propto |D|^{-m/2} \tau^{-r/2} \int \exp \left\{ \sum_{i=1}^m \sum_{j=1}^{n_i} \ell_{ij}(\beta, \psi, b_i; y_{ij}) - \frac{1}{2} \sum_{i=1}^m b_i^T D^{-1} b_i - \frac{1}{2\tau} a^T \Sigma a \right\} da db d\beta. \quad (20)$$

Let $\ell_M(\tau, \psi; y) = \log L_M(\tau, \psi; y)$ be the log-marginal likelihood function of (τ, ψ) . Zhang and Lin (2003) showed that the score $U_\tau = \partial \ell_M(\tau, \psi; y) / \partial \tau |_{\tau=0}$ can be approximated by

$$U_\tau \approx \frac{1}{2} \left\{ (Y - X\beta)^T V^{-1} N \Sigma N^T V^{-1} (Y - X\beta) - \text{tr}(P N \Sigma N^T) \right\} \Big|_{\hat{\beta}, \hat{\psi}}, \quad (21)$$

where $\hat{\beta}$ is the MLE of β and $\hat{\psi}$ is the REML estimate of ψ from the null GLMM (22), and Y is the working vector $Y = X\beta + Zb + \Delta(y - \mu)$ under the null GLMM

$$g(\mu) = X\beta + Zb, \quad (22)$$

where $\Delta = \text{diag}\{g'(\mu_{ij})\}$, $P = V^{-1} - V^{-1}X(X^T V^{-1}X)^{-1}X^T V^{-1}$ and $V = W^{-1} + Z\tilde{D}Z^T$

with $\tilde{D} = \text{diag}\{D, \dots, D\}$ and W is defined similarly as in Section 4 except μ_{ij}^0 is replaced by μ_{ij} . All these matrices are evaluated under the reduced model (22).

Write $U_\tau = \mathcal{U}_\tau - \tilde{e}$, where \mathcal{U}_τ and \tilde{e} are the first and second terms of U_τ in (21). Zhang and Lin (2003) showed that the mean of \mathcal{U}_τ is approximately equal to \tilde{e} under $H_0 : \tau = 0$. Similar to the score test derived in Section 4, the mean of \mathcal{U}_τ increases as τ increases. Therefore, we will reject $H_0 : \tau = 0$ when \mathcal{U}_τ is large, implying a one-sided test. The variance of \mathcal{U}_τ under H_0 can be approximated by

$$\tilde{I}_{\tau\tau} = I_{\tau\tau} - I_{\tau\psi}^T I_{\psi\psi}^{-1} I_{\tau\psi}, \quad (23)$$

where

$$I_{\tau\tau} = \frac{1}{2} \text{tr}\{(PN\Sigma N^T)^2\}, \quad I_{\tau\psi} = \frac{1}{2} \text{tr}\left(PN\Sigma N^T P \frac{\partial V}{\partial \psi}\right), \quad I_{\psi\psi} = \frac{1}{2} \text{tr}\left(P \frac{\partial V}{\partial \psi} P \frac{\partial V}{\partial \psi}\right). \quad (24)$$

Define $\kappa = \tilde{I}_{\tau\tau}/2\tilde{e}$ and $\nu = 2\tilde{e}^2/\tilde{I}_{\tau\tau}$. Then $S_\tau = \mathcal{U}_\tau/\kappa$ approximately has a χ_ν^2 distribution, and we will reject $H_0 : \tau = 0$ at the significance level α if $S_\tau \geq \chi_{\alpha;\nu}^2$. The simulation conducted by Zhang and Lin (2003) indicates that this modified score test for polynomial covariate effect in the semiparametric additive mixed model (16) has approximately right size and is powerful to detect alternatives.

7 Application

In this section, we illustrate the likelihood ratio testing and the score testing for variance components in GLMMs discussed in Sections 3 and 4, as well as the score polynomial covariate effect testing in GAMMs discussed in Section 6 through application to the data from Indonesian children infectious disease study (Zeger and Karim, 1991). Two hundred and seventy-five Indonesian preschool children were examined for up to six quarters for the sign of respiratory infection (0=no, 1=yes). Totally there are 2000 observation in the data set. Available covariates include: age in years; Xerophthemia status (sign for vitamin A deficiency); gender; height for age; the presence of stunning and the seasonal sine and cosine.

The primary interest of the study is to see if vitamin A deficiency has an effect on the respiratory infection adjusting for other covariates and taking into account the correlation in the data.

Zeger and Karim (1991) used Gibbs' sampling approach to fit the following logistic-normal GLMM

$$\text{logit}(P[y_{ij} = 1|b_i]) = x_{ij}^T\beta + b_i, \quad (25)$$

where y_{ij} is the respiratory infection indicator for the i th child at the j th interview, x_{ij} is the 7×1 vector of the covariates described above with corresponding effects β , $b_i \sim N(0, \theta)$ is the random effect modeling the between-child variation/between-child correlation. No statistically significant effect of vitamin A deficiency on respiratory infection was found.

We can also conduct a likelihood inference for model (25) by evaluating the required integrations using Gaussian quadrature technique. The MLE of θ is $\hat{\theta} = 0.58$ with $\text{SE}(\hat{\theta}) = 0.31$, which indicates that there may be between-child variation in the probability of getting respiratory infection. An interesting question is whether or not we can reject $H_0 : \theta = 0$. The likelihood ratio statistic for this data set is $-2\ln\lambda_m = 674.872 - 669.670 = 5.2$. The resulting p-value $= 0.5P[\chi_1^2 \geq 5.2] = 0.011$, indicating strong evidence against H_0 using the LRT procedure.

Alternatively, we may apply the score tests to test H_0 . The (corrected) score statistic for this data set is 2.678. The p-value from the one-sided score test is 0.0037, and the p-value from the two-sided score test is 0.0074. Both tests provide strong evidence against $H_0 : \theta = 0$.

Motivated by their earlier work, Zhang and Lin (2003) considered testing whether or not $f(\text{age})$ in the following semiparametric additive mixed model can be adequately represented by a quadratic function of age

$$\text{logit}(P[y_{ij} = 1|b_i]) = s_{ij}^T\beta + f(\text{age}_{ij}) + b_i, \quad (26)$$

where s_{ij} are the remaining covariates. The score test statistic described in Section 6 for $K = 3$ is $S_\tau = 5.73$ with 1.30 degrees of freedom, indicating a strong evidence against H_0 : $f(\text{age})$ is a quadratic function of age (p-value= 0.026). This may imply that nonparametric modeling of $f(\text{age})$ in model (26) is preferred.

8 Discussion

In this chapter, we have reviewed the likelihood ratio test and the score test for testing variance components in generalized linear mixed models. The central issue is that the null hypothesis usually places some of the variance components on the boundary of the model parameter space, and therefore the traditional null chi-squared distribution of the LRT statistic no longer applies and the p-value based on traditional LR chi-square distribution is often too conservative. Using the theory developed by Self and Liang (1987), we have reviewed the LRT for some special cases and show the LRT generally follows a mixture of chi-square distribution. In order to derive the right null distribution of the LRT statistic, one needs to know the (Fisher) information matrix at the true parameter value (under the null hypothesis) and the topological behavior of the neighborhood of the true parameter value. However, as our simulation indicates, the LRT for the variance components in a GLMM may suffer from numerical instability when the variance component is small and numerical integration is high dimensional.

On the other hand, the score statistic only involves parameter estimates under the null hypothesis and hence can be calculated much more straightforward and efficiently. We discussed both the one-sided score test and the much simpler two-sided score test. Both tests have the correct size. The one-sided score test has the same asymptotic distribution as the correct likelihood ratio test. Hence, similar to the LRT, the calculation of the one-sided score test requires the knowledge of the information matrix and the topological behavior of the neighborhood of the true parameter value and also requires computing a mixture of chi-

square distributions. The two-sided score test is based on the regular chi-square distribution and has the right size. It is much easier to calculate especially for testing multiple variance components. The simulation studies presented here and in the statistical literature show that the two-sided score test may suffer from some power loss compared to the (correct) likelihood ratio test and the one-sided score test.

We have also reviewed the likelihood ratio test and the score test for testing whether or not a nonparametric covariate effect in a GAMM can be adequately modeled by a polynomial of certain degree compared to a smoothing spline or a penalized spline function. Although the problem can be reduced to testing a variance component equal to zero using the mixed effects representation of the smoothing (penalized) spline, the GLMM results for likelihood ratio test and the score test for variance components do not apply because the data under mixed effects representation of the spline do not have an independent cluster structure any more. Since the LRT statistic will be prohibitive to calculate for a GLMM with potentially high dimensional random effects, we have particularly reviewed the score test of Zhang and Lin (2003) for testing the parametric covariate model versus the nonparametric covariate model in the presence of a single nonparametric covariate function. Future research is needed to develop simultaneous tests for multiple covariate effects.

REFERENCES

- Booth, J.G. and Hobert, J.P. (1999). Maximizing generalized linear mixed model likelihoods with an automated Monte Carlo EM algorithm. *Journal of the Royal Statistical Society, Series B* **61**, 265–285.
- Breslow, N.E. and Clayton, D.G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association* **88**, 9–25.
- Breslow, N.E. and Lin, X. (1995). Bias correction in generalized linear mixed models with a single component of dispersion. *Biometrika* **82**, 81–91.

- Crainiceanu, C., Ruppert, D., Claeskens, G. and Wand, M.P. (2005). Exact likelihood ratio tests for penalized splines. *Biometrika* **92**, 91-103.
- Diggle, Heagerty, Liang and Zeger (2002). *Analysis of Longitudinal Data*. Oxford University Press.
- Hall, D.B. and Praestgaard, J.T. (2001). Order-restricted score tests for homogeneity in generalized linear and nonlinear mixed models. *Biometrika* **88**, 739–751.
- Hsiao, C.K. (1997). Approximate Bayes factors when a model occurs on the boundary. *Journal of the American Statistical Association* **92**, 656–663.
- Jacqmin-Gadda, H. and Commenges, D. (1995). Tests of homogeneity for generalized linear models. *Journal of the American Statistical Association* **90**, 1237–46.
- Laird, N.M. and Ware, J.H. (1982). Random effects models for longitudinal data. *Biometrics* **38**, 963–974.
- Lin, X. (1997). Variance component testing in generalized linear models with random effects. *Biometrika* **84**, 309–326.
- Lin, X. and Breslow, N.E. (1996). Bias correction in generalized linear mixed models with multiple components of dispersion. *Journal of the American Statistical Association* **91**, 1007–1016.
- Lin, X. and Zhang, D. (1999). Inference in generalized additive mixed models using smoothing splines. *Journal of the Royal Statistical Society, Series B* **61**, 381–400.
- Self, S.G. and Liang, K.Y. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association* **82**, 605–610.

- Silvapulle, M.J. and Silvapulle, P. (1995). A score test against one-sided alternatives. *Journal of the American Statistical Association* **90**, 342-349.
- Stram, D.O. and Lee, J.W. (1994). Variance components testing in the longitudinal mixed effects model. *Biometrics*, **50**, 1171–1177.
- Verbeke, G. and Molenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data*. Springer.
- Verbeke, G. and Molenberghs, G. (2003). The use of score tests for inference on variance components. *Biometrics* **59**, 254–262.
- Verbeke, G. and Molenberghs, G. (2005). *Models for Discrete Longitudinal Data*. Springer
- Zeger, S.L. and Karim, M.R. (1991). Generalized linear models with random effects; a Gibbs sampling approach. *Journal of the American Statistical Association* **86**, 79–86.
- Zhang, D. (1997). *Generalized Additive Mixed Model*, unpublished Ph.D. dissertation, Department of Biostatistics, University of Michigan.
- Zhang, D. and Lin, X. (2003). Hypothesis testing in semiparametric additive mixed models. *Biostatistics* **4**, 57–74.